Journal of Forestry Research (2010) 21(4): 469–474
DOI 10.1007/s11676-010-0100-5

ORIGINAL PAPER

Biomass estimation of *Shorea robusta* with principal component analysis of satellite data

Nilanchal Patel • Arnab Majumdar

Received: 2010-05-03; Accepted: 2010-06-22

Abstract: Spatio-temporal assessment of the above ground biomass (AGB) is a cumbersome task due to the difficulties associated with the measurement of different tree parameters such as girth at breast height and height of trees. The present research was conducted in the campus of Birla Institute of Technology, Mesra, Ranchi, India, which is predominantly covered by Sal (*Shorea robusta* C. F. Gaertn). Two methods of regression analysis was employed to determine the potential of remote sensing parameters with the AGB measured in the field such as linear regression analysis between the AGB and the individual bands, principal components (PCs) of the bands, vegetation indices (VI), and the PCs of the VIs respectively and multiple linear regression (MLR) analysis between the AGB and all the variables in each category of data. From the linear regression analysis, it was found that only the NDVI exhibited regression coefficient value above 0.80 with the remaining parameters showing very low values. On the other hand, the MLR based analysis revealed significantly improved results as evidenced by the occurrence of very high correlation coefficient values of greater than 0.90 determined between the computed AGB from the MLR equations and field-estimated AGB thereby ascertaining their superiority in providing reliable estimates of AGB. The highest correlation coefficient of 0.99 is found with the MLR involving PCs of VIs.

Keywords: above ground biomass; spectral response modeling; vegetation indices; principal component analysis; linear and multiple regression analysis.

Introduction

Two types of remotely sensed data that have been used frequently by the researchers to extract information about biomass and stand parameters comprise the optical data and non-optical data such as Radar and Lidar data. Roy et al. (1996) employed linear and multiple regression analysis techniques to determine biomass in Madhav National Park located in Madhya Pradesh, India from the spectral reflectance values of different bands of Landsat 5 TM data and the various vegetation indices and found that significantly high correlation exists between the biomass and middle infrared band and NDVI. On the other hand, among the various tasseled cap indices, greenness index and wetness index showed positive relationship with the biomass while the brightness index was found to be negatively correlated. Dong et al. (2003) also found positive relationship of NDVI with biomass. Lu (2006) described a host of techniques such as linear and non-linear regression, k-nearest neighbour, and neural network to determine the degree of relationship between biomass and spectral data. Le´vesque (1999) used fine resolution remotely sensed data such as Airborne digital camera and found that integration of spatial transforms of image fractions (texture, semivariogram range) and raw brightness values in high-resolution remotely sensed data have a good potentiality in modeling of forest structure and health. Lefsky et al. (2001) evaluated the utility of several medium-resolution remotely sensed data for estimating stand structure attributes like age, basal area, biomass, and diameter at breast height (dbh). Their investigation revealed that multitemporal TM should be investigated as an alternative to either hyperspectral or hyperspatial sensors, which are more expensive and more difficult to process than multitemporal Landsat TM. Only a few researches have been conducted regarding the biomass estimation using the coarse resolution data such as NOAA Advanced Very High Resolution Radiometer (AVHRR), SPOT VEGETATION, and Moderate Resolution Imaging Spectro Radiometer (MODIS). However, coarse resolution data could be significantly useful for rapid estimation of biomass at continental and global scales (Fraser et al. 2002).

The present study was carried out for the following objectives: To determine the coefficient of regression between the above ground biomass measured from the field with the spectral reflectance in individual bands of IRS LISS III data, individual principal components generated from the different spectral bands,

The online version is available at http://www.springerlink.com

Nilanchal Patel (✉) • Arnab Majumdar
Birla Institute of Technology Mesra, Ranchi – 835215, Jharkhand, India
Phone Nos.: 0091-651-2276003 (O), Email: npatelrsgis@yahoo.com

Responsible editor: Chai Ruihai

seven individual vegetation indices (Table 1) were considered in the present study, and the individual principal components generated from seven vegetation indices respectively employing linear regression analysis; To develop multiple linear regression equations between the above ground biomass measured from the field with the spectral reflectance in different bands of IRS LISS III data, principal components generated from the different spectral bands, seven vegetation indices, and the principal components generated from seven vegetation indices respectively employing multiple linear regression analysis; To determine the correlation coefficients between the above ground biomass computed through the multiple linear regression equations developed for four different types of remotely sensed parameters and the above ground biomass measured from the field in order to assess to what extent the various remote sensing parameters could provide reliable estimates of the field-estimated above ground biomass.

Materials and Methods

Study area

The study area selected for performing the present research covers the campus of the Birla Institute of Technology Mesra, Ranchi located within the Ranchi city, the capital of Jharkhand, India (Fig. 1) and is covered in the Survey of India toposheet No. 73 E/7. The study area represents natural and protected open type forest with Sal as the dominant species. The study area is surrounded by Subarnarekha and Jumar River. There are three main seasons that prevail in the region: summer season (15 March - 15 June) when the climate is extreme with maximum temperature of 38°C and minimum 21°C, rainy season (June - October) with the annual rainfall being about 1530 mm and the winter season (November - February) with temperature ranging between 2°C and 22°C. The vegetation remains green during the rainy and winter season while the summer season is characterized by both leaf shedding and leaf emerging phases. The soil type within the study area is largely lateritic (Kumar 1999).

Materials used

To perform the present investigation, LISS III data of IRS-P6 satellite (Path/Row- 105/55) acquired on 20th October, 2008 was used. The LISS III data comprises four spectral bands viz. Green (0.52–0.59 µm), Red (0.62–0.68 µm), NIR (0.77–0.86 µm) and MIR (1.55-1.70µm) each having 23.5m spatial resolution and 8-bit radiometric resolution. Erdas Imagine 9.1® was employed for performing all digital image processing tasks such as georeferencing of the raw image, subsetting of the plots, extraction of reflectance values, generation of vegetation indices and principal component images. The ARCMAP 9.2® was used to generate canopy density and species type maps. The geographic coordinates of the different sampling plots selected within the study area were determined using Leica Handheld Global Positioning System. The linear regression and multiple regression analyses between the biomass and the other governing parameters were performed using the SPSS 17.0 software. MS-Excel software was used for generation of plots between the biomass and vegetation parameters and remote sensing variables such as reflectance values in different bands, vegetation indices and PCs of bands and PCs of vegetation indices. The original satellite image was georeferenced with the GCPs extracted from the Survey of India toposheet No. 73E/7 on 1:50 000 scale.

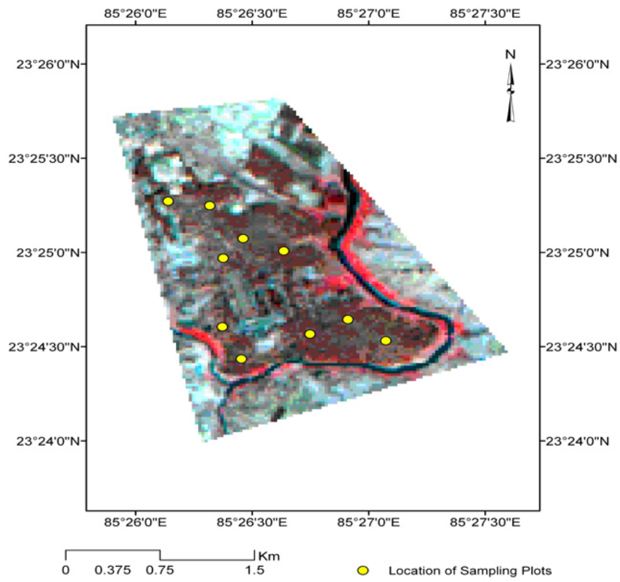


Fig. 1 Study area showing the locations of sampling plots

Methods used

The task of biomass estimation of the Sal trees through remote sensing involved two broad steps as mentioned below.

Methods used for field data collection and above ground biomass estimation

Collection of various field data pertaining to the Sal trees and subsequently, their above ground biomass estimation involved the following steps.

Keeping in view the extent of the study area and the variation in the canopy density of the Sal trees within the study area, ten representative sampling plots were selected. At the time of selecting the plots it was kept in mind that they should be large enough to ensure that there was no influence by edges, road, and/or pipelines.

The size of each sampling plot was taken as 1 hectare, which is approximately equal to 16 pixels of the LISS III data. Exact geographic coordinates of these plots in the field were determined using GPS (Leica Handheld Global Positioning System).

Total number of trees existing within each sampling plot was counted.

The gbh (girth at breast height) of the individual trees occurring within each plot was measured using measuring tape and was later converted to dbh (diameter at breast height) by dividing

the gbh values by π (0.318).

An important parameter for the volume estimation of trees through volumetric equation is the height of tree. The measurement of the tree height was carried out through indirect approach. First, using a hypsometer the viewing angle between the observer and position at the first forcking of the tree was measured and then the distance between the observer and the base of the tree was measured using a measuring tape. The height of the tree was determined using the following formula.

$$\text{Tan}\ \theta = \text{height of the tree / distance from tree} \quad (1)$$

where θ denotes the viewing angle.

The volume of each tree was determined using the following formula,

$$V = 0.118 + 0.257 D^2 * H \quad (2)$$

where, V is the volume of Sal trees, D is the diameter at breast height, and H is the height at the first forcking.

The total above ground biomass of a plot is calculated as the sum of the total bole biomass of all the trees occurring within the plot and the litter biomass contained in the plot. The bole biomass represents the dry weight of the standing tree whereas the litter biomass comprises the dry weight of the leaves lying on the ground. The bole biomass of each tree was determined by multiplying the volume of the tree with its specific gravity. For Sal trees the specific gravity has been determined as 0.726 (Shrestha 2003). The litter biomass within each plot was determined using the following procedure. The litter lying within a few small sampling plots of size (1m×1m) selected at different locations within the main plot were collected from the ground and then their dry weight was measured in the laboratory which was then averaged to obtain the litter biomass of a 1×1m plot. The total litter biomass of the main plot was determined by multiplying the average litter biomass corresponding to 1×1m plot with the area of the main plot. Then the total bole biomass and total litter biomass estimated for each main plot was summed up to determine the total above ground biomass within each main plot.

Methods employed for estimation of above ground biomass using the parameters derived from remotely sensed data

The satellite data was georeferenced with the help of the Survey of India toposheet No. 73 E/7 on 1:50 000 scale with sub-pixel accuracy. Then each sampling plot selected in the field was correctly delineated on the satellite image using the coordinates measured in the field through GPS.

Different vegetation indices such as NDVI, SAVI, RVI, DVI, IPVI, GNDVI and GVI (Table 1) were generated through model maker tool in ERDAS IMAGINE 9.1 software. This was followed by the generation of principal components of the bands and stacked vegetation indices respectively.

Table 1: Equations of different vegetation indices used in this study

Vegetation Indices	Explanation of symbols	References
$\text{NDVI} = \frac{(\text{NIR} - \text{R})}{(\text{NIR} + \text{R})}$	NDVI=Normalized Difference Vegetation Index, NIR = near infrared, R = red, G = green, and L = Soil-adjustment factor	Rouse et al. (1974)
$\text{RVI} = \frac{\text{NIR}}{\text{R}}$	RVI (Ratio Vegetation Index)	Richardson and Wiegand (1977)
$\text{SAVI} = \frac{(\text{NIR} - \text{R})}{(\text{NIR} + \text{R} + \text{L})} + (1 + \text{L})$	SAVI (Soil-adjusted Vegetation Index)	Huete (1988)
$\text{IPVI} = \frac{\text{NIR}}{(\text{NIR} + \text{R})}$	IPVI (Infrared Percentage Vegetation Index)	Crippen (1990)
$\text{GNDVI} = \frac{(\text{NIR} - \text{G})}{(\text{NIR} + \text{G})}$	GNDVI (Green Normalized Vegetation Index)	Jensen (2006)
$\text{GVI} = \frac{(\text{NIR} - \text{R})}{(\text{NIR} + \text{R}) \times 127} + 128$	GVI (Green Vegetation Index)	Jensen (2006)
$\text{DVI} = (\text{NIR} - \text{R})$	DVI (Difference Vegetation Index)	Lillesand et al. (2007)

Then the average value of the reflectance in different bands, different vegetation indices, individual PCs generated from the bands and stacked vegetation indices for each sampling plot were computed.

Both linear and multiple linear regression analyses were performed between the average AGB value and different remote sensing parameters as mentioned in the preceding step for the individual plots. Regression plots were also generated between the AGB and individual remote sensing parameters. Multiple linear regression analyses were performed employing the SPSS 17.0 software.

Multiple linear regression equations were developed between the field-estimated AGB for each of the ten different plots selected in the entire study area and the corresponding remote sensing based parameters viz. the four spectral bands of LISS III data, four principal components generated from the spectral bands, seven vegetation indices considered in the present study, and the seven principal components determined from the seven vegetation indices respectively. Then AGB was computed for each sampling plot by substituting the corresponding reflectance values of four spectral bands, PC values of the spectral bands, seven vegetation indices and the PC values of the seven vegetation indices in the respective multiple linear regression equations. Correlation analysis was performed between the field-estimated

AGB and the AGB estimated through the multiple linear regression analysis for different sampling plots to determine to what extent the regression based AGB tallies with the field estimated AGB and consequently, to determine to what extent the remotely sensed parameters could provide correct estimates of the AGB measured in the field.

Finally, comparative analysis was performed among the results obtained from the various regression analyses in order to determine the particular spectral band, vegetation index, and PC(s) of bands and vegetation indices that exhibited the highest regression coefficient with the field-estimated AGB.

Results

The results determined from the present research have been presented in two sections; section one comprises the discussion about the linear regression analysis performed between AGB and individual remote sensing parameters and section two deals with the multiple linear regression analysis.

Linear Regression Analysis

As mentioned earlier, linear regression analysis was performed between the field-estimated AGB determined from each of the ten sampling plots and the corresponding remotely sensed parameters such as the average reflectance values recorded in different spectral bands, individual vegetation indices, individual PCs generated from the spectral bands and the vegetation indices respectively (Table 2). Comparative analysis of the regression coefficient values (R^2) given in table indicates the occurrence of significantly higher R^2 values with the vegetation indices and their PCs as compared to that exhibited by the different spectral bands and their PCs. Among the four spectral bands of LISS III data used in the present study, Red band showed the highest regression coefficient with the field-estimated AGB with R^2 value of 0.65 while the remaining three bands exhibited nearly same R^2 values (Fig. 2). Similarly, the highest value of regression coefficient (0.66) is exhibited by PC2 while the remaining three PCs show nearly similar R^2 values ranging between 0.56 and 0.59 (Fig. 3). On the other hand, the different vegetation indices exhibit significantly much higher regression coefficients with the AGB as compared to the spectral bands and the PCs of the spectral bands. Among the different vegetation indices used in the study, NDVI shows the highest regression coefficient (0.81) followed in decreasing order by SAVI (R^2 = 0.78), IPVI (R^2 = 0.75), GVI (R^2 = 0.73), RVI (R^2 = 0.71), DVI (R^2 = 0.70) and GNDVI (R^2 = 0.66) (Fig. 4). The PCs of the seven vegetation indices also exhibit considerably high regression coefficient with the AGB which is found to be lower than that exhibited by the vegetation indices but significantly higher than the regression coefficient values exhibited by the spectral bands and the PCs of the spectral bands respectively. The first PC of the vegetation indices exhibits the highest regression coefficient with a value of 0.77 followed in decreasing order by PC2 (R^2 = 0.72), PC3 (R^2 = 0.67), PC4 (R^2 = 0.64), PC6 (R^2 = 0.55), PC5 (R^2 = 0.54) and PC7 (R^2 = 0.52) respectively (Fig. 5).

Table 2: Regression coefficient values of different remote sensing parameters with field-estimated AGB

Remote Sensing Parameters	Individual Parameters	R^2
Spectral Bands	Green	0.60
	Red	0.65
	NIR	0.62
	MIR	0.59
PCs of Spectral Bands	PC 1	0.57
	PC 2	0.66
	PC 3	0.59
	PC 4	0.56
Vegetation Indices	NDVI	0.81
	SAVI	0.78
	RVI	0.71
	DVI	0.70
	IPVI	0.75
	GNDVI	0.66
	GVI	0.73
PCs of Vegetation Indices	PC 1	0.77
	PC 2	0.72
	PC 3	0.67
	PC 4	0.64
	PC 5	0.54
	PC 6	0.55
	PC 7	0.52

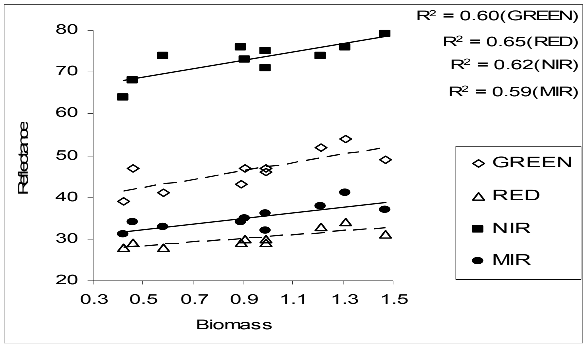


Fig. 2 Relationships between AGB and spectral bands

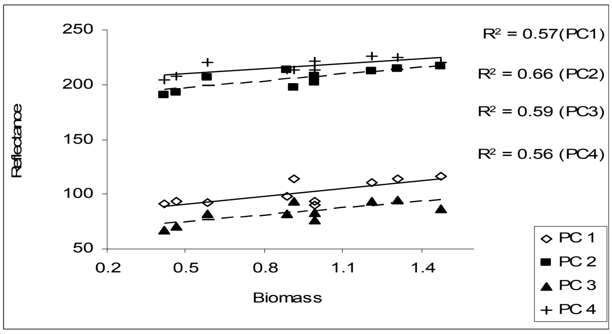


Fig. 3 Relationships between AGB and PCs of spectral bands

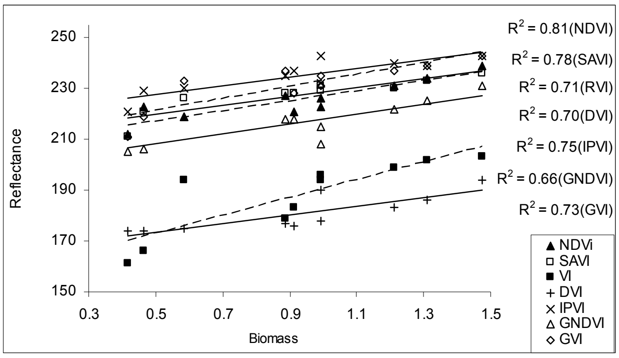


Fig. 4 Relationships between AGB and vegetation indices

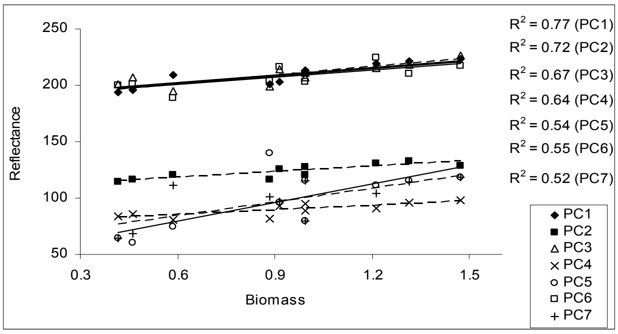


Fig. 5 Relationships between AGB and PCs of vegetation indices

From the foregoing analyses it can be inferred that the field-estimated AGB is best regressed with the NDVI values while the SAVI also exhibited significantly high regression coefficient with the AGB. The other remote sensing parameters such as the spectral bands, PCs of the spectral bands, remaining five vegetation indices and PCs of the seven vegetation indices failed to provide considerable regression coefficients with the AGB. Therefore, NDVI and SAVI can be considered as reliable estimators of the AGB (Majumdar 2010).

Multiple Linear Regression Analysis

Multiple linear regressions were developed between the AGB estimated for each of the ten different sampling plots and the corresponding average DN values of each of the four remote sensing based parameters considered in the present study viz. the spectral bands, PCs of the spectral bands, seven vegetation indices, and the PCs of the seven vegetation indices and are shown in equations (3), (4), (5) and (6) respectively.

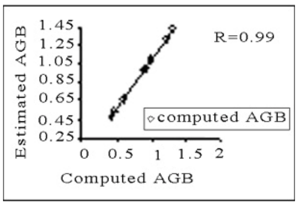
$$AGB_{Bands} = 0.006^* Green + 0.051^* Red + 0.043^* NIR + 0.020^* MIR - 4.707 \quad (3)$$

$$AGB_{PCs\ of\ Bands} = 0.018^* PC\ 1 + 0.019^* PC\ 2 - 0.007^* PC\ 3 + 0.009^* PC\ 4 - 6.218 \quad (4)$$

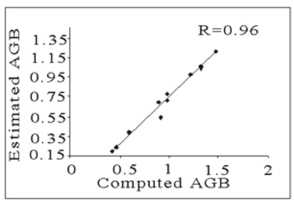
$$AGB_{Vegetation\ Indices} = 0.024^* NDVI - 0.027^* SAVI + 0.009^* RVI + 0.004^* DVI + 0.024^* IPVI + 0.001^* GNDVI + 0.004^* GVI - 7.663 \quad (5)$$

$$AGB_{PCs\ of\ Vegetation\ Indices} = 0.015^* PC\ 1 + 0.002^* PC\ 2 + 0.007^* PC\ 3 + 0.011^* PC\ 4 + 0.006^* PC\ 5 + 0.000^* PC\ 6 - 0.001^* PC\ 7 - 5.295 \quad (6)$$

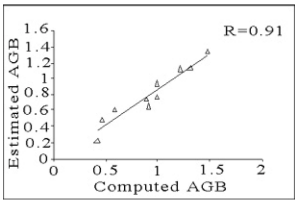
Subsequently, the AGB of each sampling plot was computed by substituting the corresponding average DN values of the different remote sensing parameters viz. the spectral bands, PCs of the spectral bands, seven vegetation indices, and PCs of the vegetation indices in the respective multiple linear regression equations. This is referred to as the computed AGB of the different sampling plots. The primary objective of the present research is to assess the potential of different remote sensing derived parameters for providing the correct estimates of the AGB measured from the field. In order to perform this task, the correlation coefficient values between the computed AGB derived from different multiple linear regressions and the AGB estimated from the field were determined (Figs. 6a, 6b, 6c, 6d). Comparative analysis of these figures revealed the occurrence of the highest correlation coefficient of 0.99 between the computed AGB based on the principal components of the vegetation indices and the field-estimated AGB (Fig. 6a), followed in decreasing order by the vegetation indices (0.96), spectral bands (0.92) and principal components of the spectral bands (0.91) (Figs. 6b, 6c, 6d respectively). These findings prove that the multiple linear regression derived on the basis of the PCs of the seven vegetation indices selected for the present investigation is capable of providing estimates of the field-measured AGB up to 100 percent accuracy while the multiple linear regressions derived based on the remaining three remote sensing parameters could also be considered as potential reliable estimators of the above ground biomass. The occurrence of the very high correlation coefficients of the computed AGB with the field-estimated AGB based on the multiple linear regressions using all the four types of remote sensing data could be attributed to the combined contribution of the spectral information from the multivariable data and also, the prevalence of monotonous vegetation (Sal) with nearly homogeneous canopy density thereby resulting in similar spectral characteristics throughout the study area.



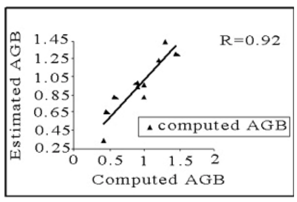
a: PCs of Vegetation Indices



b: Vegetation Indices



c: Spectral Bands



d: PCs of Spectral Bands

Fig. 6 Correlation between computed AGB and field-estimated AGB

Discussions

The above ground biomass estimation from the remote sensing data performed in the present study involved two types of regression analysis such as the linear regression analysis and multiple linear regression analysis. Based on the results of the linear regression analysis performed between the above ground biomass and different remote sensing parameters such as the individual spectral bands of IRS LISS III data, individual PCs generated from the spectral bands, seven vegetation indices, and the individual principal components generated from the seven vegetation indices, it was found that it is only the NDVI that exhibited a regression coefficient value of greater than 0.80 with the above ground biomass. In addition, the PC1 of the vegetation indices showed a regression coefficient of 0.72. The other remote sensing parameters exhibited significantly very low regression coefficient values with the AGB.

On the other hand, the multiple linear regression analysis performed between the field-estimated above ground biomass and various types of remote sensing parameters such as the spectral bands, PCs of the spectral bands, vegetation indices and PCs of the vegetation indices exhibited significantly higher values of regression coefficients. The correlation coefficient values determined between the computed above ground biomass derived from the multiple linear regression equations and the field-estimated above ground biomass were found to be significantly very high; for example, 0.99 for the principal components of the vegetation indices, 0.96 for the vegetation indices, 0.92 for the spectral bands and 0.91 for the principal components of the spectral bands. The occurrence of significantly high values of correlation coefficients can be attributed to the combined contribution of the spectral information of the vegetation canopy from the multiple variables. Two salient findings that emerged from the present investigation are as follows. First, the multiple linear regression based analysis possesses better potential of providing correct estimates of the above ground biomass as compared to the linear regression based analysis and second, the principal components of the vegetation indices are better correlated with the above ground biomass.

Acknowledgments
The authors express their sincere thanks to the Vice Chancellor of the Birla Institute of Technology, Mesra, Ranchi, India for providing the necessary facilities to carry out the desired research in the Digital Image Processing and GIS laboratory of the Department of Remote Sensing. This research comprises a part of the M.Sc. dissertation of Mr. Arnab Majumdar under the supervision of Prof. (Dr.) Nilanchal Patel.

References

Crippen RE. 1990. Calculating the vegetation index faster. *Remote Sensing of Environment*, **34**: 71–73.

Dong J, Kaufmann RK, Myneni RB, Tucker CJ, Kauppi PE, Liski J, Buermann W, Alexeyev V, Hughes MK. 2003. Remote sensing estimates of boreal and temperate forest woody biomass: carbon pools, sources, and sinks. *Remote Sensing of Environment*, **84**: 393–410.

Fraser RH, Li Z. 2002. Estimating fire-related parameters in boreal forest using SPOT VEGETATION. *Remote Sensing of Environment*, **82**: 95–110.

Huete AR. 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment,* **25**(3): 295–309.

Jensen JR. 2006 (2nd Edition). *Remote Sensing of the Environment: An Earth Resource Perspective.* South Carolina: Prentice Hall, 592pp.

Kumar R. 1999. Characterization of Rice Growing Ecosystem Using Remote Sensing and GIS Techniques. Ph.D. Thesis (unpublished), Birsa Agricultural University, Ranchi, India.

Lefsky MA, Cohen WB., Spies TA. 2001. An evaluation of alternate remote sensing products for forest inventory, monitoring, and mapping of Douglas-fir forests in western Oregon. *Canadian Journal of Forest Research*, **31**(1): 78–87.

Levesque J, King DJ. 1999. Airborne digital camera image semivariance for evaluation of forest structural damage at an acid mine site. *Remote Sensing of Environment*, **68**: 112–124.

Lillesand TM, Kiefer RW, Chipman JW. 2007 (6th Edition). *Remote Sensing and Image Interpretation*. Madison: John Wiley & Sons, 784pp.

Lu D. 2006. The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing*, vol. **27**(7): 1297–1328.

Majumdar A. 2010. Biomass Estimation of Sal (*Shorea robusta*) through Principal Component Analysis of Satellite Data. M.Sc. Thesis (unpublished), Birla Institute of Technology Mesra, Ranchi, India.

Richardson AJ, Wiegand CL. 1977. Distinguishing vegetation from soil background information. *Photogramnetric Engineering and Remote Sensing*, **43(**12):1541–1552.

Rouse JW Jr, Haas RH, Deering DW, Schell JA, Harlan JC. 1974. Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation. NASA/GSFC Type III Final Report, Greenbelt, MD: p. 371.

Roy PS, Ravan SA. 1996. Biomass estimation using satellite remote sensing data - An investigation on possible approaches for natural forest. *Journal of Biosciences*, **21**(4): 535–561.

Shrestha HL. 2003. Comparative Evaluation of Different Spatial Resolution Data for Timber Volume Estimation. M.Sc. Thesis, International Institute for Geoinformation Science and Earth Observation, Enschede, Netherlands.